Richa Rambhia
December 12, 2021

Module 5 R Practice
Report & Outputs

Introduction

The task in this week is to examine and test the relationships between the variables of the dataset using the regression model and produce the correlation and regression table for the same. The dataset chosen for this task is the Fish dataset which has a record of 7 common different fish species in fish market sales along with the attributes like weight, height, width, and length of the fish. With the help of the regression model on this dataset, we can predict the weight of the fish based on its various attributes.

The attributes of the dataset are as follows:

Species: Species name of the fish

Weight: Weight of the fish in Gram g

Length 1: Vertical length in cm

Length 2: Diagonal length in cm

Length 3: Cross length in cm

Height: Height in cm

Width: Diagonal width in cm

This dataset and its attributes will therefore be used in order to understand the relationship between the variables and to predict the weight of the fish. Descriptive analysis and data visualization would be performed to understand the data before starting with the predictive analysis. Finally, the correlation and regression analysis would be done which will in better understanding of the data.

Source to the dataset: https://www.kaggle.com/aungpyaeap/fish-market

Data Analysis

For the regression and correlation analysis, packages like 'broom', 'corrplot', and 'gtsummary' was installed and imported. After the libraries were installed and imported, the dataset was read into a variable named 'dataset_fish' and this would be the dataset which will be used throughout in this task. Once the dataset was read into a variable, it was necessary to have a clear understanding of the data and so the dataset description was done for the same. Descriptive analysis is where we understand the statistical values of the attributes to have some idea about the variables and so for the attributes like weight, width and height, descriptive analysis was performed. The outputs of the above-mentioned points are as follows.

Output:

1. Reading the csv file

```
Console  Terminal ×  Jobs ×
R 3.6.3 · ~/
> #reading csv file
> dataset_fish <- read.csv("Fish.csv")
> dataset_fish
   ï..Species Weight Length1 Length2 Length3  Height  Width
1       Bream  242.0    23.2    25.4    30.0 11.5200 4.0200
2       Bream  290.0    24.0    26.3    31.2 12.4800 4.3056
3       Bream  340.0    23.9    26.5    31.1 12.3778 4.6961
4       Bream  363.0    26.3    29.0    33.5 12.7300 4.4555
5       Bream  430.0    26.5    29.0    34.0 12.4440 5.1340
6       Bream  450.0    26.8    29.7    34.7 13.6024 4.9274
7       Bream  500.0    26.8    29.7    34.5 14.1795 5.2785
8       Bream  390.0    27.6    30.0    35.0 12.6700 4.6900
9       Bream  450.0    27.6    30.0    35.1 14.0049 4.8438
10      Bream  500.0    28.5    30.7    36.2 14.2266 4.9594
11      Bream  475.0    28.4    31.0    36.2 14.2628 5.1042
12      Bream  500.0    28.7    31.0    36.2 14.3714 4.8146
13      Bream  500.0    29.1    31.5    36.4 13.7592 4.3680
14      Bream  340.0    29.5    32.0    37.3 13.9129 5.0728
15      Bream  600.0    29.4    32.0    37.2 14.9544 5.1708
16      Bream  600.0    29.4    32.0    37.2 15.4380 5.5800
17      Bream  700.0    30.4    33.0    38.3 14.8604 5.2854
18      Bream  700.0    30.4    33.0    38.5 14.9380 5.1975
19      Bream  610.0    30.9    33.5    38.6 15.6330 5.1338
20      Bream  650.0    31.0    33.5    38.7 14.4738 5.7276
21      Bream  575.0    31.3    34.0    39.5 15.1285 5.5695
22      Bream  685.0    31.4    34.0    39.2 15.9936 5.3704
23      Bream  620.0    31.5    34.5    39.7 15.5227 5.2801
24      Bream  680.0    31.8    35.0    40.6 15.4686 6.1306
25      Bream  700.0    31.9    35.0    40.5 16.2405 5.5890
26      Bream  725.0    31.8    35.0    40.9 16.3600 6.0532
27      Bream  720.0    32.0    35.0    40.6 16.3618 6.0900
28      Bream  714.0    32.7    36.0    41.5 16.5170 5.8515
29      Bream  850.0    32.8    36.0    41.6 16.8896 6.1984
30      Bream 1000.0    33.5    37.0    42.6 18.9570 6.6030
31      Bream  920.0    35.0    38.5    44.1 18.0369 6.3063
32      Bream  955.0    35.0    38.5    44.0 18.0840 6.2920
33      Bream  925.0    36.2    39.5    45.3 18.7542 6.7497
34      Bream  975.0    37.4    41.0    45.9 18.6354 6.7473
35      Bream  950.0    38.0    41.0    46.5 17.6235 6.3705
36      Roach   40.0    12.9    14.1    16.2  4.1472 2.2680
37      Roach   69.0    16.5    18.2    20.3  5.2983 2.8217
38      Roach   78.0    17.5    18.8    21.2  5.5756 2.9044
39      Roach   87.0    18.2    19.8    22.2  5.6166 3.1746
40      Roach  120.0    18.6    20.0    22.2  6.2160 3.5742
41      Roach    0.0    19.0    20.5    22.8  6.4752 3.3516
42      Roach  110.0    19.1    20.8    23.1  6.1677 3.3957
43      Roach  120.0    19.4    21.0    23.7  6.1146 3.2943
44      Roach  150.0    20.4    22.0    24.7  5.8045 3.7544
45      Roach  145.0    20.5    22.0    24.3  6.6339 3.5478
46      Roach  160.0    20.5    22.5    25.3  7.0334 3.8203
47      Roach  140.0    21.0    22.5    25.0  6.5500 3.3250
48      Roach  160.0    21.1    22.5    25.0  6.4000 3.8000
49      Roach  169.0    22.0    24.0    27.2  7.5344 3.8352
50      Roach  161.0    22.0    23.4    26.7  6.9153 3.6312
51      Roach  200.0    22.1    23.5    26.8  7.3968 4.1272
52      Roach  180.0    23.6    25.2    27.9  7.0866 3.9060
53      Roach  290.0    24.0    26.0    29.2  8.8768 4.4968
```

2. Describing the data

```
> summary_dataset <- summary(dataset_fish)
> summary_dataset
    ï..Species     Weight           Length1         Length2         Length3          Height      
 Bream    :35   Min.   :   0.0   Min.   : 7.50   Min.   : 8.40   Min.   : 8.80   Min.   : 1.728  
 Parkki   :11   1st Qu.: 120.0   1st Qu.:19.05   1st Qu.:21.00   1st Qu.:23.15   1st Qu.: 5.945  
 Perch    :56   Median : 273.0   Median :25.20   Median :27.30   Median :29.40   Median : 7.786  
 Pike     :17   Mean   : 398.3   Mean   :26.25   Mean   :28.42   Mean   :31.23   Mean   : 8.971  
 Roach    :20   3rd Qu.: 650.0   3rd Qu.:32.70   3rd Qu.:35.50   3rd Qu.:39.65   3rd Qu.:12.366  
 Smelt    :14   Max.   :1650.0   Max.   :59.00   Max.   :63.40   Max.   :68.00   Max.   :18.957  
 Whitefish: 6                                                                                    
     Width      
 Min.   :1.048  
 1st Qu.:3.386  
 Median :4.248  
 Mean   :4.417  
 3rd Qu.:5.585  
 Max.   :8.142  
                

> 
> classtype <- sapply(dataset_fish, class)
> classtype
ï..Species     Weight    Length1    Length2    Length3     Height      Width 
  "factor"  "numeric"  "numeric"  "numeric"  "numeric"  "numeric"  "numeric" 
> 
```

```
Console   Terminal   Jobs
R 3.6.3 · ~/
> #describing the dataset
> colnames(dataset_fish)
[1] "ï..Species" "Weight"     "Length1"    "Length2"    "Length3"    "Height"     "Width"
>
> start_records <- head(dataset_fish,10)
> start_records
   ï..Species Weight Length1 Length2 Length3  Height  Width
1       Bream    242    23.2    25.4    30.0 11.5200 4.0200
2       Bream    290    24.0    26.3    31.2 12.4800 4.3056
3       Bream    340    23.9    26.5    31.1 12.3778 4.6961
4       Bream    363    26.3    29.0    33.5 12.7300 4.4555
5       Bream    430    26.5    29.0    34.0 12.4440 5.1340
6       Bream    450    26.8    29.7    34.7 13.6024 4.9274
7       Bream    500    26.8    29.7    34.5 14.1795 5.2785
8       Bream    390    27.6    30.0    35.0 12.6700 4.6900
9       Bream    450    27.6    30.0    35.1 14.0049 4.8438
10      Bream    500    28.5    30.7    36.2 14.2266 4.9594
>
> end_records <- tail(dataset_fish,10)
> end_records
    ï..Species Weight Length1 Length2 Length3 Height  Width
150      Smelt    9.8    10.7    11.2    12.4 2.0832 1.2772
151      Smelt    8.7    10.8    11.3    12.6 1.9782 1.2852
152      Smelt   10.0    11.3    11.8    13.1 2.2139 1.2838
153      Smelt    9.9    11.3    11.8    13.1 2.2139 1.1659
154      Smelt    9.8    11.4    12.0    13.2 2.2044 1.1484
155      Smelt   12.2    11.5    12.2    13.4 2.0904 1.3936
156      Smelt   13.4    11.7    12.4    13.5 2.4300 1.2690
157      Smelt   12.2    12.1    13.0    13.8 2.2770 1.2558
158      Smelt   19.7    13.2    14.3    15.2 2.8728 2.0672
159      Smelt   19.9    13.8    15.0    16.2 2.9322 1.8792
>
> dim(dataset_fish)
[1] 159   7
>
> str(dataset_fish)
'data.frame':   159 obs. of  7 variables:
 $ ï..Species: Factor w/ 7 levels "Bream","Parkki",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Weight    : num  242 290 340 363 430 450 500 390 450 500 ...
 $ Length1   : num  23.2 24 23.9 26.3 26.5 26.8 26.8 27.6 27.6 28.5 ...
 $ Length2   : num  25.4 26.3 26.5 29 29 29.7 29.7 30 30 30.7 ...
 $ Length3   : num  30 31.2 31.1 33.5 34 34.7 34.5 35 35.1 36.2 ...
 $ Height    : num  11.5 12.5 12.4 12.7 12.4 ...
 $ Width     : num  4.02 4.31 4.7 4.46 5.13 ...
>
```

3. Descriptive analysis

```
Console   Terminal   Jobs
R 3.6.3 · ~/
> #descriptive analysis
>
> #weight
> min(dataset_fish$Weight)
[1] 0
> max(dataset_fish$Weight)
[1] 1650
> mean(dataset_fish$Weight)
[1] 398.3264
> median(dataset_fish$Weight)
[1] 273
> mode(dataset_fish$Weight)
[1] "numeric"
> sd(dataset_fish$Weight)
[1] 357.9783
> range(dataset_fish$Weight)
[1]    0 1650
>
> #height
> min(dataset_fish$Height)
[1] 1.7284
> max(dataset_fish$Height)
[1] 18.957
> mean(dataset_fish$Height)
[1] 8.970994
> median(dataset_fish$Height)
[1] 7.786
> mode(dataset_fish$Height)
[1] "numeric"
> sd(dataset_fish$Height)
[1] 4.286208
> range(dataset_fish$Height)
[1]  1.7284 18.9570
>
> #width
> min(dataset_fish$Width)
[1] 1.0476
> max(dataset_fish$Width)
[1] 8.142
> mean(dataset_fish$Width)
[1] 4.417486
> median(dataset_fish$Width)
[1] 4.2485
> mode(dataset_fish$Width)
[1] "numeric"
> sd(dataset_fish$Width)
[1] 1.685804
> range(dataset_fish$Width)
[1] 1.0476 8.1420
>
```

The frequency table was created to understand the unique values and the count of the species of the fish in the dataset. Also, a subset of the dataset was created which had the all the numeric attributes except for the species column for the correlation table and chart generated at a later stage.

Output:

```
> #frequency table
> unique_species <- unique(dataset_fish$ï..Species)
> unique_species
[1] Bream     Roach     Whitefish Parkki    Perch     Pike      Smelt
Levels: Bream Parkki Perch Pike Roach Smelt Whitefish
>
> count_species <- count(dataset_fish$ï..Species)
> count_species
          x freq
1     Bream   35
2    Parkki   11
3     Perch   56
4      Pike   17
5     Roach   20
6     Smelt   14
7 Whitefish    6
>
> table_species <- table(dataset_fish$ï..Species)
> table_species

    Bream    Parkki     Perch      Pike     Roach     Smelt Whitefish
       35        11        56        17        20        14         6
>
```

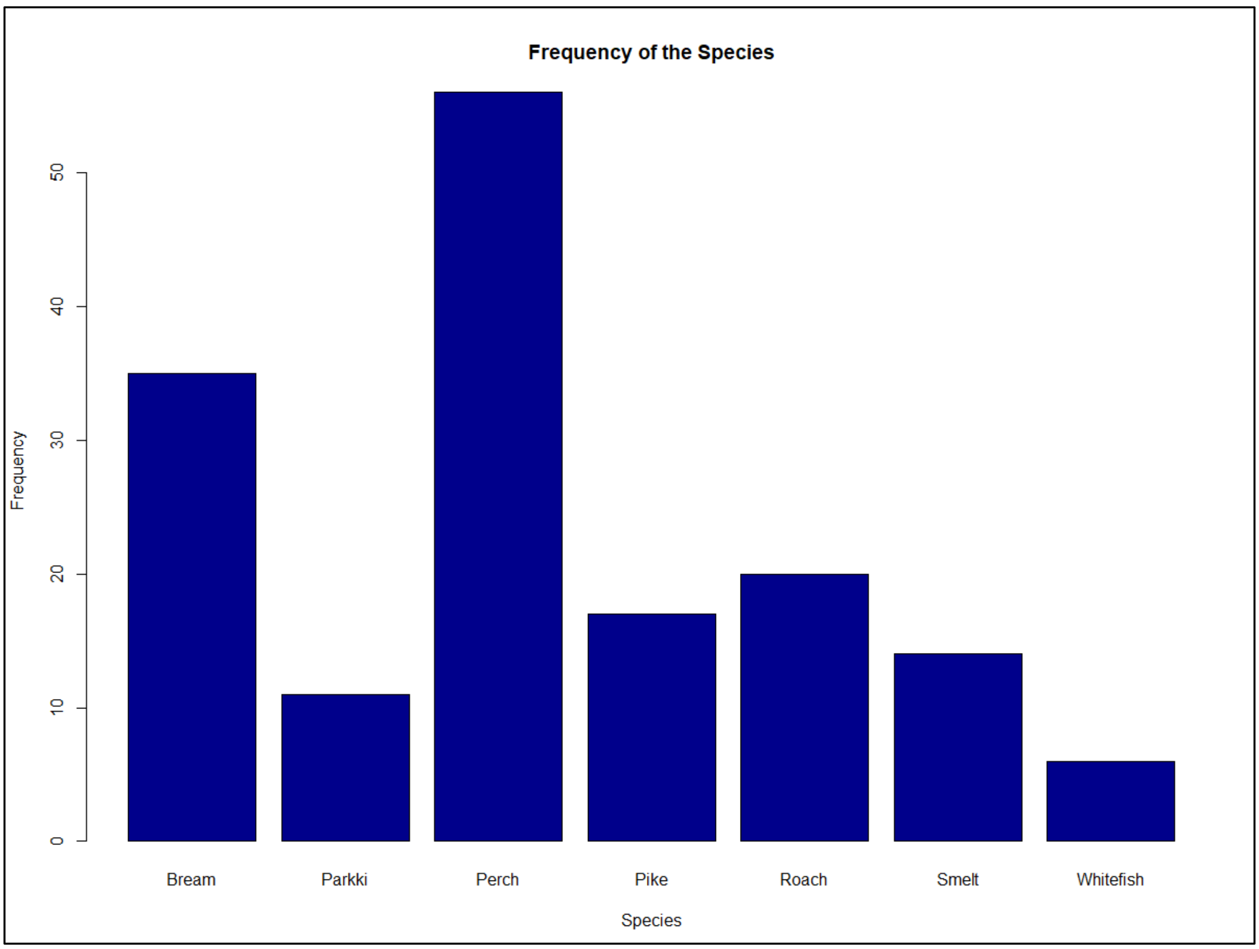
```
> #creating new subset of dataset
> dataset_fish_new <- data.frame(dataset_fish$Weight, dataset_fish$Length1, dataset_fish$Length2, dataset_fis
h$Length3,
+                                dataset_fish$Height, dataset_fish$Width)
> dataset_fish_new
    dataset_fish.Weight dataset_fish.Length1 dataset_fish.Length2 dataset_fish.Length3
1                 242.0                 23.2                 25.4                 30.0
2                 290.0                 24.0                 26.3                 31.2
3                 340.0                 23.9                 26.5                 31.1
4                 363.0                 26.3                 29.0                 33.5
5                 430.0                 26.5                 29.0                 34.0
6                 450.0                 26.8                 29.7                 34.7
7                 500.0                 26.8                 29.7                 34.5
8                 390.0                 27.6                 30.0                 35.0
9                 450.0                 27.6                 30.0                 35.1
10                500.0                 28.5                 30.7                 36.2
11                475.0                 28.4                 31.0                 36.2
12                500.0                 28.7                 31.0                 36.2
13                500.0                 29.1                 31.5                 36.4
14                340.0                 29.5                 32.0                 37.3
15                600.0                 29.4                 32.0                 37.2
16                600.0                 29.4                 32.0                 37.2
17                700.0                 30.4                 33.0                 38.3
18                700.0                 30.4                 33.0                 38.5
19                610.0                 30.9                 33.5                 38.6
20                650.0                 31.0                 33.5                 38.7
21                575.0                 31.3                 34.0                 39.5
22                685.0                 31.4                 34.0                 39.2
23                620.0                 31.5                 34.5                 39.7
24                680.0                 31.8                 35.0                 40.6
25                700.0                 31.9                 35.0                 40.5
26                725.0                 31.8                 35.0                 40.9
27                720.0                 32.0                 35.0                 40.6
28                714.0                 32.7                 36.0                 41.5
29                850.0                 32.8                 36.0                 41.6
30               1000.0                 33.5                 37.0                 42.6
31                920.0                 35.0                 38.5                 44.1
32                955.0                 35.0                 38.5                 44.0
33                925.0                 36.2                 39.5                 45.3
34                975.0                 37.4                 41.0                 45.9
35                950.0                 38.0                 41.0                 46.5
36                 40.0                 12.9                 14.1                 16.2
37                 69.0                 16.5                 18.2                 20.3
38                 78.0                 17.5                 18.8                 21.2
39                 87.0                 18.2                 19.8                 22.2
40                120.0                 18.6                 20.0                 22.2
41                  0.0                 19.0                 20.5                 22.8
42                110.0                 19.1                 20.8                 23.1
```

Data Visualization

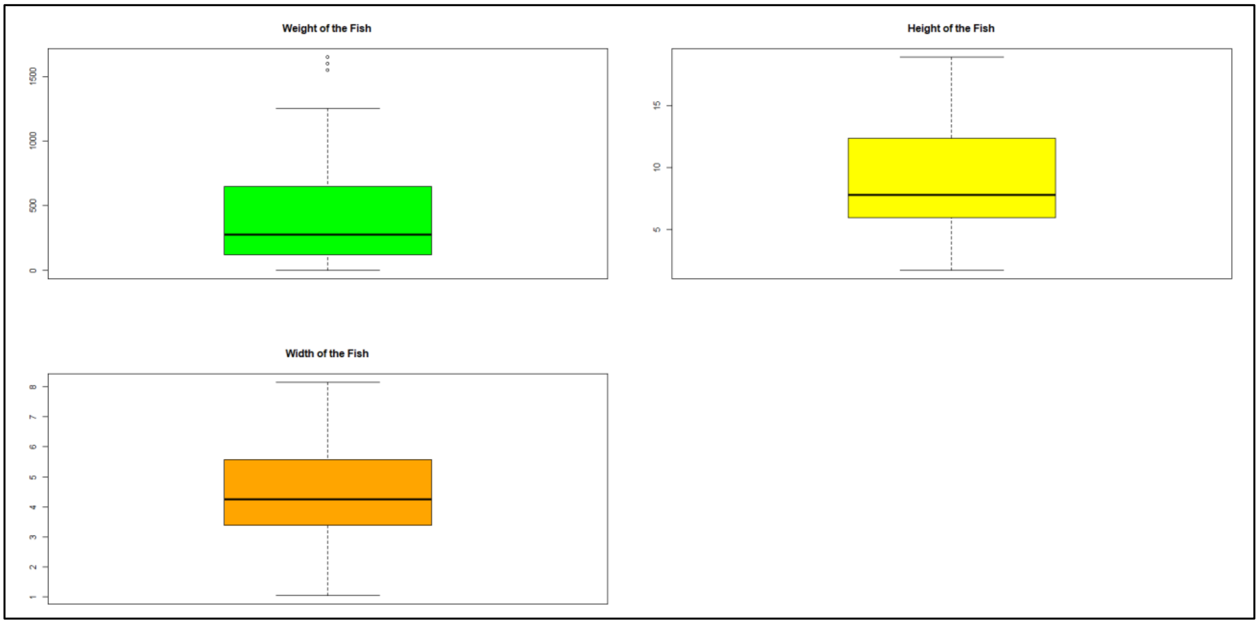
The next step was data visualizations which gave a better understanding of the dataset. The graphs created based on the dataset were frequency of the species, boxplot of the attributes, histogram of the attributes and density plot of the attributes. Frequency barplot graph helped in understanding the frequency of the species and its unique values. Boxplot gave an idea about the statistical values of the attributes like weight, width, and height of the fish. Histogram for the three attributes were generated to help understand the frequency and count value of the attributes. And finally, density plot for the attributes weight, width, and height were also created. The outputs of the visualizations created are as follows.

Output:

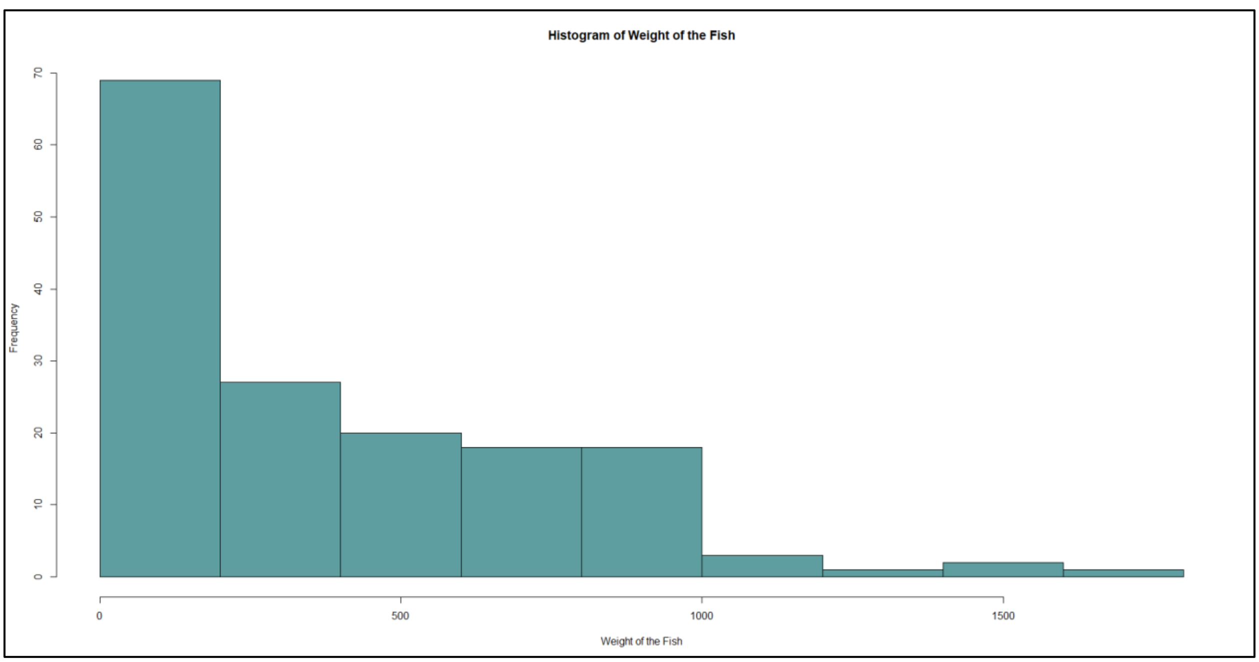
1. Graph 1: Frequency of the Species



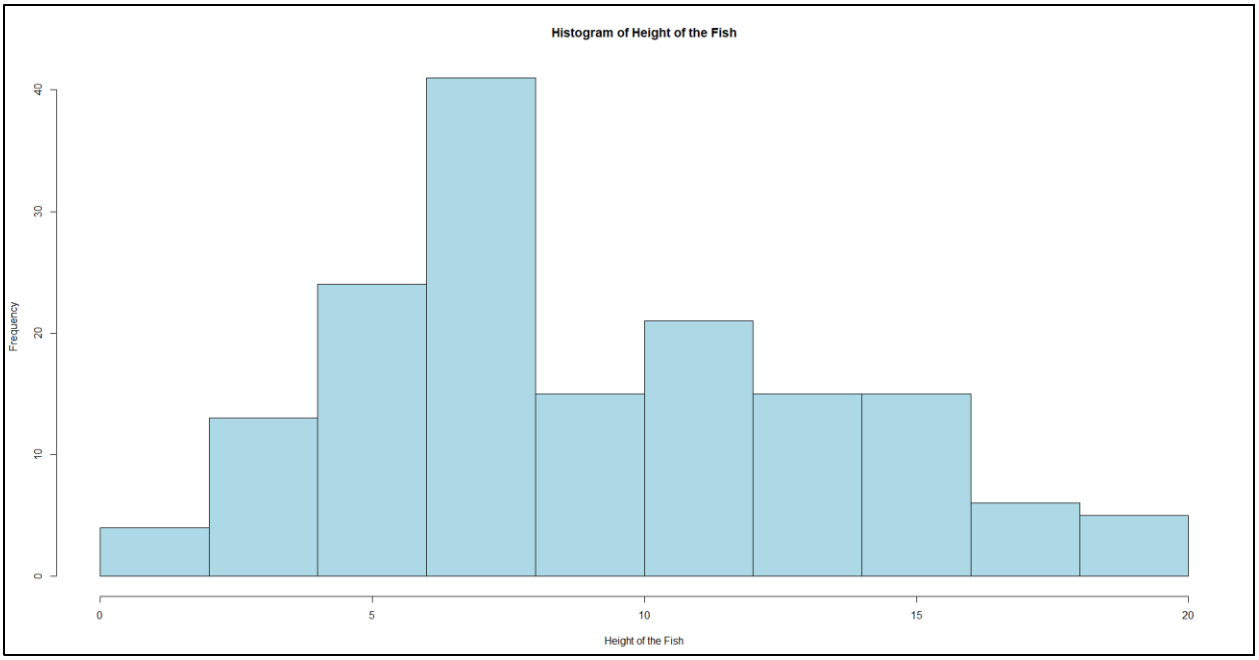
2. Graph 2: Boxplot of the attributes



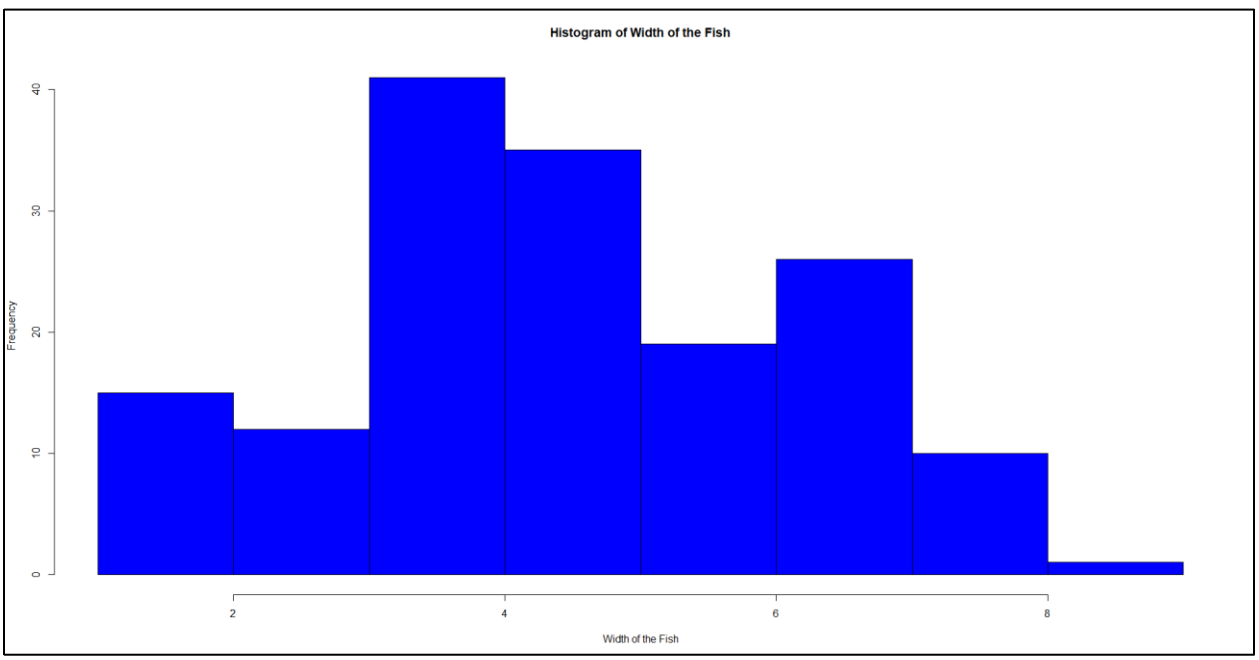
3. Graph 3: Histogram of Weight of the Fish



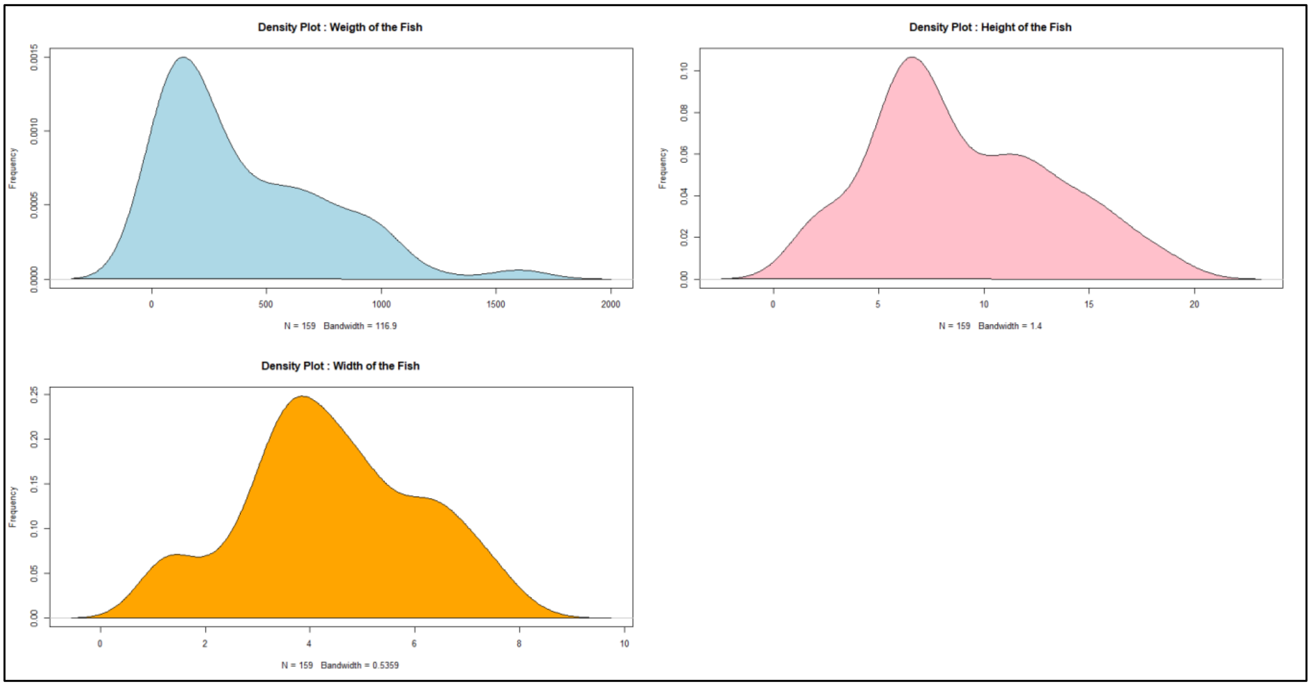
4. Graph 4: Histogram of Height of the Fish



5. Graph 5: Histogram of Width of the Fish



6. Graph 6: Density plot of the attributes



Regression Model

Regression model is a statistical model which estimates the relationship between one dependent variable and one or more independent variables using a line called as the regression line. The relationship between the variables is described using the models by fitting a line to the observed data. The independent variables are who do not depend on any other variables and are independently controlled inputs whereas the dependent variables are whose changes solely depends on another variable which is the independent variable, and which means that the value of the dependent variable changes only if the independent variable changes.

Correlation coefficient is a statistical concept which helps in establishing a relation between the predicted and actual values of the dataset and the calculated value of the correlation coefficient explains the exactness between the predicted and actual values. The value of the correlation coefficient always lies between -1 to +1 which means that if the value of the correlation coefficient is positive then there is a similar relation between the two variables otherwise it states the dissimilarity between the two variables.

Here, to examine the relationship between the two variables and to understand the correlation coefficient, regression model was built to determine the relationship between the weight and height of the fish and the weight and width of the fish which would help in predicting the weight of the fish. At first, we examined the relationship between the weight and height of the fish where we computed the correlation coefficient value and generated a linear regression model for the same along with the plot of the regression line. Similar analysis was performed to examine the relationship between the weight and width of the fish. The outputs for the above-mentioned points are as follows.

Output & Analysis:

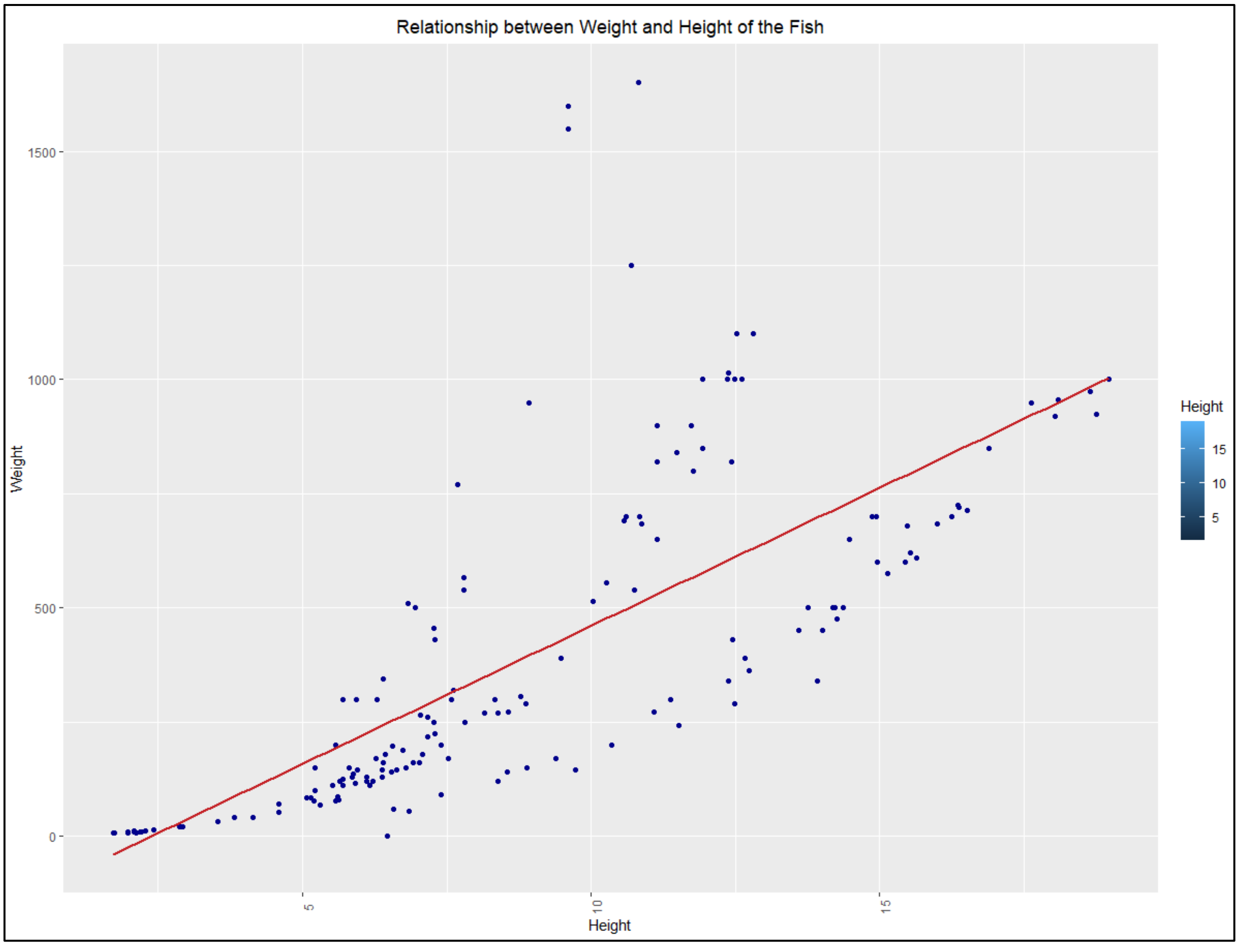
1. Relationship between Weight and Height of the Fish
    a. Correlation Coefficient Value

```
Console  Terminal ×  Jobs ×
R 3.6.3 · ~/
> #relationship between weight and height of the fish
>
> #correlation
> correlation_value1 <- cor(dataset_fish$Weight, dataset_fish$Height)
> correlation_value1
[1] 0.7243453
> |
```

    b. Regression Model

```
Console  Terminal ×  Jobs ×
R 3.6.3 · ~/
> #linear regression model & plot
> linearregression_model1 <- lm(Weight ~ Height, data = dataset_fish)
> linearregression_model1

Call:
lm(formula = Weight ~ Height, data = dataset_fish)

Coefficients:
(Intercept)       Height
     -144.4         60.5

> summary(linearregression_model1)

Call:
lm(formula = Weight ~ Height, data = dataset_fish)

Residuals:
    Min      1Q  Median      3Q     Max
-357.29 -116.53  -70.76   33.91 1163.62

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -144.386     45.663  -3.162  0.00188 **
Height        60.496      4.595  13.164  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 247.6 on 157 degrees of freedom
Multiple R-squared:  0.5247,	Adjusted R-squared:  0.5216
F-statistic: 173.3 on 1 and 157 DF,  p-value: < 2.2e-16

> |
```

c. Regression Line Plot



Here, the dependent variable is the weight of the fish, and the independent variable is the height of the fish. From the outputs, it is observed that the value of the correlation coefficient is positive which means that there is a similar relation between the two variables i.e., the weight and height of the fish. The regression model states the coefficients which helps in predicting the weight of the fish. Here, the coefficients values are -144.4 and 60.5 and the prediction equation for the weight would be: *Weight = -144.4 + 60.5 x Height*. This equation thus would help us in predicting the weight of the fish. The effect here is approx. 60.5 which means that as the height of the fish increases the weight of the fish increases by 60.5 units. Also, since the p-value is less than 0.05, we reject the null hypothesis which tells us that the height of the fish is not significant enough to determine the weight of the fish as there are other factors and parameters involved that helps in determining the weight of the fish. Lastly, a regression line is plotted for height of the fish against the weight of the fish and the line passing through the points is called the fitted line which tells us about the relationship between the two variables

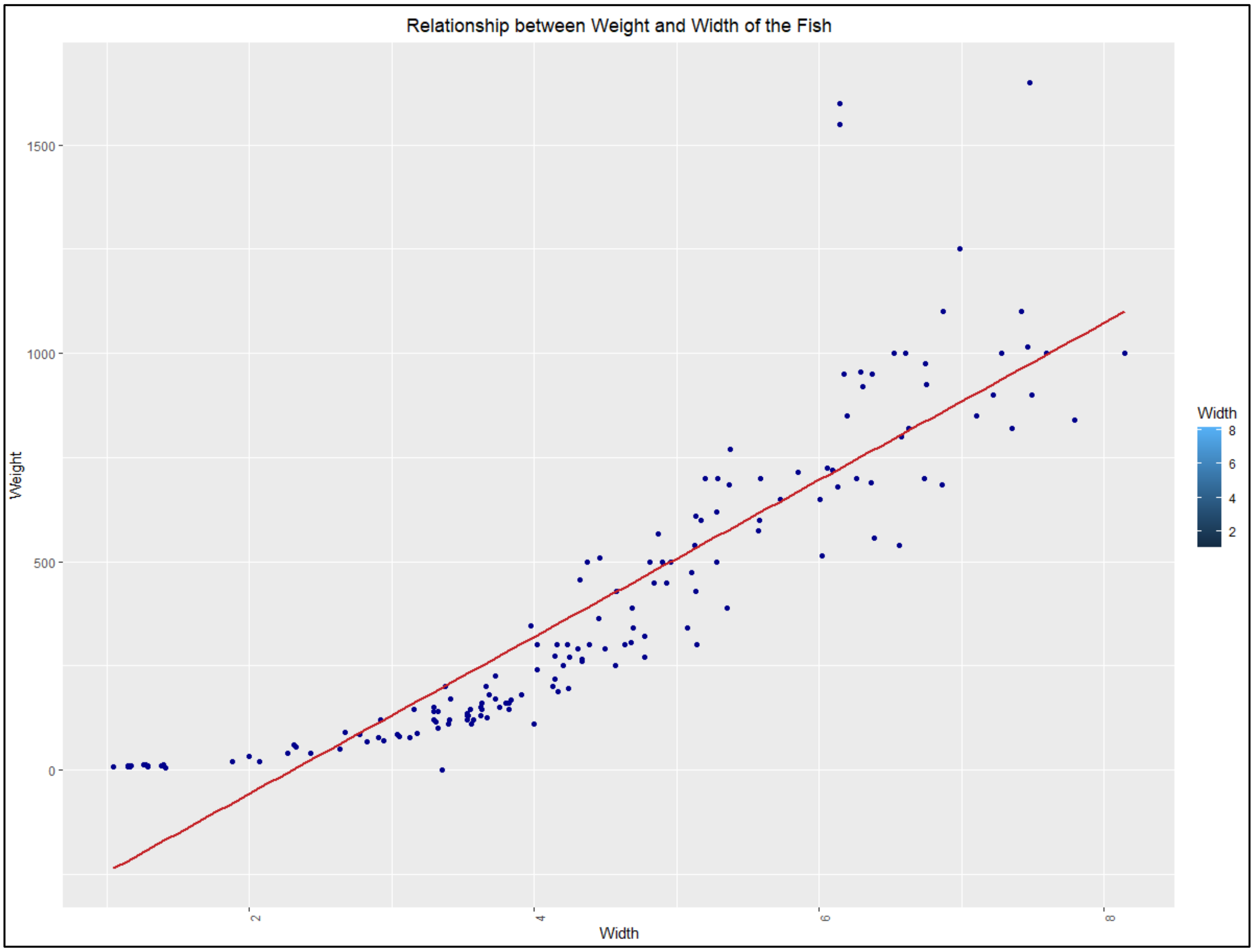
2. Relationship between Weight and Width of the Fish
    a. Correlation Coefficient Value

```
Console   Terminal ×   Jobs ×
R 3.6.3 · ~/
> #relationship between weight and width of the fish
>
> #correlation
> correlation_value2 <- cor(dataset_fish$Weight, dataset_fish$Width)
> correlation_value2
[1] 0.8865066
>
```

    b. Regression Model

```
Console   Terminal ×   Jobs ×
R 3.6.3 · ~/
> #linear regression model & plot
> linearregression_model2 <- lm(Weight ~ Width, data = dataset_fish)
> linearregression_model2

Call:
lm(formula = Weight ~ Width, data = dataset_fish)

Coefficients:
(Intercept)        Width
     -433.3        188.2

> summary(linearregression_model2)

Call:
lm(formula = Weight ~ Width, data = dataset_fish)

Residuals:
    Min      1Q  Median      3Q     Max
-262.03 -101.16  -42.59   69.68  876.66

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -433.259     37.063  -11.69   <2e-16 ***
Width        188.249      7.842   24.01   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 166.2 on 157 degrees of freedom
Multiple R-squared:  0.7859,	Adjusted R-squared:  0.7845
F-statistic: 576.3 on 1 and 157 DF,  p-value: < 2.2e-16

>
```

c. Regression Line Plot



In this case, the dependent variable is again the weight of the fish but the independent variable this time is the width of the fish. From the above outputs, we observe here that the value of the correlation coefficient is positive which states that there is a similar relation between the two variables which in this case is the weight and width of the fish. In the regression model built, it is observed that the coefficient values are -433.259 and 188.249 and the prediction equation is: *Weight = -433.259 + 188.249 x Width* where this equation helps in predicting the weight of the fish for the independent variable, width. The effect here is approx. 188.2 which tells us that the weight of the fish increases by 188.2 units as the width of the fish increases. Now, since the p-value is less than 0.05, we reject the null hypothesis and that implies that the width of the fish is not only the significant attribute to determine the weight of the fish as we have other parameters involved too which helps in predicting the weight of the fish. Lastly, a regression line is plotted for width of the fish against the weight of the fish and the line passing through the points is called the fitted line which tells us about the relationship between the two variables.

Correlation Table & Chart

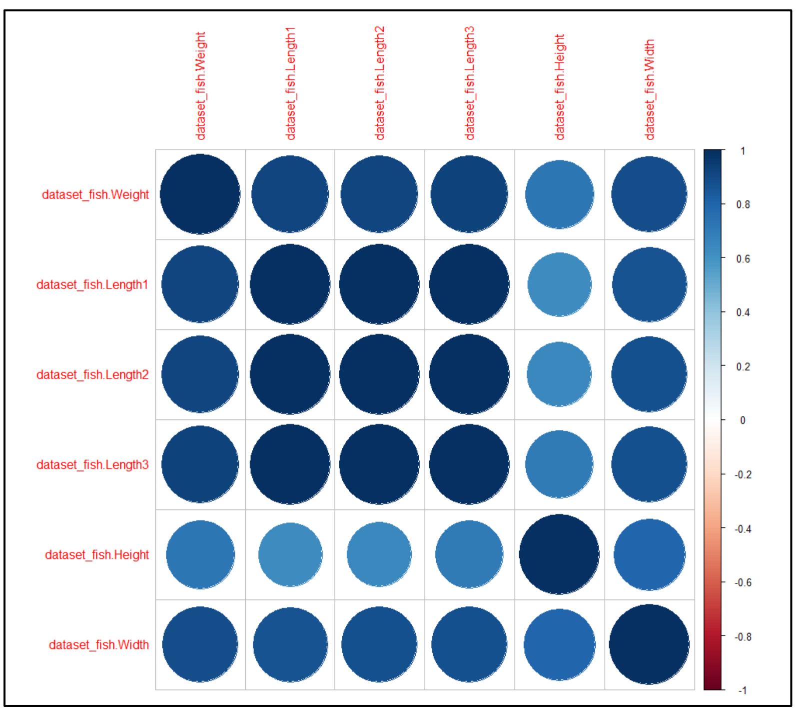
A correlation table shows the correlation coefficients between the variables where each cell in the table shows the correlation between the two variables. This table is used to summarize data as an input into a more advanced analysis. The correlation chart on the other hand helps in understanding the relationship between the two variables by visual analysis. The output of the correlation table and the correlation chart for the fish dataset is as below.

Output:

a. Correlation Table:

```
Console   Terminal ×   Jobs ×
R 3.6.3 · ~/
> #correlation table
> dataset_fish_new.cor = cor(dataset_fish_new)
> dataset_fish_new.cor
                      dataset_fish.Weight dataset_fish.Length1 dataset_fish.Length2 dataset_fish.Length3
dataset_fish.Weight             1.0000000            0.9157117            0.9186177            0.9230436
dataset_fish.Length1            0.9157117            1.0000000            0.9995173            0.9920310
dataset_fish.Length2            0.9186177            0.9995173            1.0000000            0.9941026
dataset_fish.Length3            0.9230436            0.9920310            0.9941026            1.0000000
dataset_fish.Height             0.7243453            0.6253779            0.6404408            0.7034089
dataset_fish.Width              0.8865066            0.8670497            0.8735467            0.8785202
                      dataset_fish.Height dataset_fish.Width
dataset_fish.Weight             0.7243453          0.8865066
dataset_fish.Length1            0.6253779          0.8670497
dataset_fish.Length2            0.6404408          0.8735467
dataset_fish.Length3            0.7034089          0.8785202
dataset_fish.Height             1.0000000          0.7928810
dataset_fish.Width              0.7928810          1.0000000
>
```

b. Correlation Chart:



Regression Table

A regression table gives us information about the statistical values of the regression model built for the particular dataset like the characteristic variable, beta value, 95% confidence interval and the p-value. The output of the regression table created for the two models are as below.

Output:

a. Regression Table for Model 1

Characteristic	Beta	95% CI[1]	p-value
Height	60	51, 70	<0.001

[1] CI = Confidence Interval

b. Regression Table for Model 2

Characteristic	Beta	95% CI[1]	p-value
Width	188	173, 204	<0.001

[1] CI = Confidence Interval

Summary

Regression models helps in examining the relationship between the two variables of the dataset where the correlation coefficient is a statistical concept which would help in determining the relationship between the predicted value and the actual value. Correlation and Regression are both terms used in statistics which are related to each other but are not the same. Correlation analysis basically measures the linear association between the two variables x and y which has a value between -1 and 1 where -1 indicates a perfectly negative linear correlation between two variables, 0 indicates no linear correlation between the two variables, and 1 indicates a perfectly positive linear correlation between the variables. On the other hand, Regression analysis is a method which is used to understand that how changing the values of x variable affects the values of the y variable and so in a regression model we have one variable, x, as the predictor variable and the other variable, y, as the response variable.

The task in this assignment was to examine the relationship between the two variables of the dataset using the regression model. After the dataset was read and descriptive analysis was performed on the dataset, data visualization was performed to have a better understanding of the data for further analysis. In the data visualizations, boxplot, barplot, and density plot for the different attributes of the dataset were plotted. The frequency of the species of the fish was also plotted to understand the count of each species in the dataset. Once we were clear on the dataset part, the next phase was to build a model to understand the relationship between the two variables. Here, two regression models were built, one for the relationship between the weight and height of the fish and other for the relationship between the weight and width of the fish. The correlation coefficient was also computed which in both the cases came out to be positive indicating that there is a similar relation between the two variables.

The regression model 1 examined the relationship between the weight and height of the fish where the equation of the line was, Weight = -144.4 + 60.5 x Height, where y=weight, c=-144.4, x=height, and m=60.5 which is the slope. Similarly, the model 2 examined the relationship between the weight and width of the fish where the equation of line was, Weight = -433.259 + 188.249 x Width, where y=weight, c=-433.259, x=width, and m=188.249.

From the analysis of the regression model and summary, it was observed that in both the cases the p-value was less than 0.05 which is why we rejected the null hypothesis which states that the weight of the fish does not necessarily depend on one factor or parameter, there are multiple other attributes that contribute to the prediction of the weight of the fish.

References

Prabhakaran, S. (n.d.). Linear Regression. R-Statistics.Co. Retrieved December 7, 2021, from http://r-statistics.co/Linear-Regression.html#:~:text=The%20most%20common%20metrics%20to%20look%20at%20while,zero%20the%20better%20%207%20more%20rows%20

Linear Regression in R. (n.d.). Datacamp. Retrieved December 7, 2021, from https://www.datacamp.com/community/tutorials/linear-regression-R

Zach, S. (2021, February 1). Correlation vs. Regression: What’s the Difference? Statology.Org. Retrieved December 8, 2021, from https://www.statology.org/correlation-vs-regression/#:~:text=Differences%3A%20Regression%20is%20able%20to%20show%20a%20cause-and-effect,variable%2C%20based%20on%20the%20value%20of%20another%20variable.

Tutorial: tbl_regression. (2020, September 13). Danieldsjoberg.Com. Retrieved December 8, 2021, from https://www.danieldsjoberg.com/gtsummary/articles/tbl_regression.html#:~:text=The%20tbl_regression%20%28%29%20function%20takes%20a%20regression%20model,creates%20highly%20customizable%20analytic%20tables%20with%20sensible%20defaults.

Facer, C. (n.d.). How to Create a Correlation Matrix in R. DisplayR. Retrieved December 8, 2021, from https://www.displayr.com/how-to-create-a-correlation-matrix-in-r/